Research Article

JOURNAL OF CHEMOMETRICS

Received: 7 July 2008, Revised: 18 November 2008, Accepted: 20 November 2008, Published online in Wiley Interscience: 24 February 2009

(www.interscience.wiley.com) DOI: 10.1002/cem.1212

Exponential data fitting using multilinear algebra: the decimative case

Jean-Michel Papy[a]*, Lieven De Lathauwer[b] and Sabine Van Huffel[b]

This paper presents a high-precision method for the estimation of the parameters of a signal, modelled as a finite sum of complex damped exponentials, whose poles may be close. It is well known that increasing the sampling rate increases the accuracy of traditional methods, like total least squares (TLSs)-ESPRIT and HTLS, in the case of closely spaced exponentials. However, in the case of long sequences, the computational complexity makes these methods hard to apply. To overcome this problem, the HTLSDstack and the HTLSDsum methods make use of decimated sequences. In this paper, we propose a higher order counterpart of HTLSDstack, based on concepts from multilinear algebra. Basically, it consists of decimating an oversampled signal by a factor *D*, making sure that no aliasing occurs. Each sequence is arranged in a Hankel matrix and the *D* Hankel matrices are stacked in a third-order tensor. A dimensionality reduction algorithm is applied to the tensor and the resulting subspace is used to retrieve the parameters of interest. The dimensionality reduction is possibly unsymmetric in the different modes of the tensor, which allows us to take into account ill conditioning when poles are close. The new method yields better results than HTLSDstack while its complexity stays fairly low compared to non-decimative methods.
Copyright © 2009 John Wiley & Sons, Ltd.

Keywords: complex damped exponential; harmonic retrieval; parameter estimation; decimation; Vandermonde decomposition; multilinear algebra; higher-order tensor; Tucker compression; parallel factor analysis

1. INTRODUCTION

In various applications of digital signal processing, such as nuclear magnetic resonance (NMR) spectroscopy [1–7], audio processing [8–11], speech processing [12,13], the shape from moments problem [14] or material health monitoring [15] and complex damped exponentials are used as a model function. The discrete-time signal x_n is represented as a sum of K discrete-time complex damped exponentials, contaminated by circularly symmetric white Gaussian noise (WGN) b_n:

$$x_n = \sum_{k=1}^{K} a_k \exp\{j\varphi_k\} \exp\{(-\alpha_k + j\omega_k)n\Delta t\} + b_n \qquad n = 0, \ldots, N-1 \tag{1}$$

where $j = \sqrt[2]{-1}$, n is the time index, and Δt, which is the inverse of the sampling rate ν_s, is the sampling time interval. The parameters of this model are the amplitudes a_k, the phases φ_k, the damping factors α_k and the pulsations ω_k. The frequencies ν_k can be obtained from the pulsations by means of the equality $\omega_k = 2\pi\nu_k$. Model (1) is often written in a more compact form

$$x_n = \sum_{k=1}^{K} c_k z_k^n + b_n \quad n = 0, \ldots, N-1 \tag{2}$$

where $c_k = a_k \exp\{j\varphi_k\}$ denotes the kth complex amplitude including the phase. $z_k = \exp\{(-\alpha_k + j\omega_k)\Delta t\}$ is called the kth pole of the signal.

In the last decades, various techniques have been derived in order to estimate the signal poles z_k from the data (2). This problem is referred to as harmonic retrieval. Among the most important contributions, one can cite maximum likelihood methods [16,17], linear prediction methods [18], constraint nonlinear minimization methods [19], subspace-based methods [20–22], matrix pencil techniques [23] and higher order statistics-based methods [24]. Our work focuses on subspace-based methods, which originate from Pisarenko's method [25]. In his work, Pisarenko used the eigenvalue decomposition (EVD) of the sample covariance matrix in order to determine the signal subspace, which contains the relevant signal information. A famous and widely used subspace-based algorithm is ESPRIT, which exploits the rotational invariance of signal subspaces spanned by two temporally shifted data sequences. A first version of the ESPRIT algorithm was based on the least squares (LS) solution [26] and a second one on the total least squares (TLS) solution [27]. ESPRIT also uses the EVD of the sample covariance matrix. As far as the complex exponential model is concerned, there is an exact equivalence between the EVD of the sample

* *Correspondence to: J.-M. Papy, Flanders' MECHATRONICS Technology Centre (FMTC), Celestijnenlaan 300D, bus 4027, Leuven 3001, Belgium.*
E-mail: jean-michel.papy@fmtc.be

a *J.-M. Papy*
Flanders' MECHATRONICS Technology Centre (FMTC), Celestijnenlaan 300D, bus 4027, Leuven 3001, Belgium

b *L. De Lathauwer, S. Van Huffel*
Department of Electrical Engineering (ESAT), Katholieke Universiteit Leuven, SCD-SISTA, Belgium

Copyright © 2009 John Wiley & Sons, Ltd.

covariance matrix and the singular value decomposition (SVD) of a Hankel data matrix. However, generally the SVD provides more numerical stability than the EVD. The data matrix equivalent of ESPRIT is known, in the literature on magnetic resonance spectroscopy (MRS), as HSVD [28] and the data matrix equivalent of TLS-ESPRIT is known as HTLS [3,4,29]. Although it is suboptimal [30], the HTLS algorithm has proven to be robust, reliable and easy to implement. Basically, it consists of arranging the N samples of the signal x_n in a $(L \times M)$ Hankel matrix $\boldsymbol{H}$ with $L + M - 1 = N$, $L > K$ and $M > K$. Then, after computation of the (ordered) SVD of $\boldsymbol{H}$, the K dominant left singular vectors are used to retrieve the model parameters.

An important issue is the behaviour of these methods when two (or more) signal poles are very close to each other. As any other method, HTLS performs poorly when applied to closely spaced sinusoids. To get more accurate estimates, one may increase the sampling rate. However, as the sampling rate increases, the computational cost of the HTLS algorithm gets quickly prohibitive due to the large size of the data matrix $\boldsymbol{H}$. In References [31,32], three decimative versions of the HTLS method are derived, among which HTLSDstack turns out to be the most interesting method because it has approximately the same accuracy as HTLS while having a much lower computational cost. This is the data matrix counterpart of the decimative TLS-ESPRIT algorithm presented in References [33,34]. In HTLSDstack the signal is decimated by a factor D. The D downsampled sequences are arranged in separate Hankel matrices. Then, these Hankel matrices are stacked in a block Hankel matrix of which the SVD is computed. The rest of the algorithm is similar to the HTLS algorithm. In HTLSDstack the downsampled sequences can be considered as different channels that allow for an improved frequency separation. This algorithm is, in essence, similar to the HTLSDstack algorithm, which applies to a set of time series (channels) [35,36]. Using the HTLSDstack algorithm, we are able to 'zoom in' very accurately into a spectral region D times narrower than the original one and centred around 0 Hz.

It has been proven that in the multi-channel case, multilinear algebra applies naturally and generally improves the results [37]. The problem of harmonic analysis was for the first time cast in a tensor framework in References [38,39]. In Reference [37] a multi-channel exponential data modelling technique based on multilinear algebra was derived. This method can be seen as the higher order counterpart of HTLSDstack. Instead of stacking the different Hankel matrices in a block Hankel matrix, they are stacked in a three-way array or third-order tensor. A variant of the technique for delayed bursts of exponentials was presented in Reference [40]. In this paper, we show that we can arrange decimated data in a three-way array, which leads to a higher order version of HTLSDstack having a higher robustness.

The main contributions of this paper are:

- We map an oversampled sum of complex exponentials to a third-order Hankel-type tensor. To this end, the original tensor is artificially interpreted as a series of channels, each corresponding to a decimated subsequence. Each matrix slice of the tensor is a Hankel matrix representation of such a decimated sequence. The tensor admits a third-order Vandermonde Decomposition that better captures the data structure than decompositions of the block Hankel matrix used in HTLSDstack.
- We derive a tensor-based subspace algorithm for the extraction of the signal poles. We start from the results presented in Reference [37].
- The algorithm in Reference [37] is based on the computation of the best multilinear rank-(R_1, R_2, R_3) approximation of the data tensor, where R_1, R_2 and R_3 are the theoretical 1-mode, 2-mode and 3-mode rank, respectively (see further). This approximation allows one to determine the dominant R_1-dimensional subspace of the column space of the noisy tensor. In this paper, we derive the fundamental result that the very same subspace can be estimated from the best multilinear rank-$(R_1, R_2, \tilde{R}_3)$ approximation of the data tensor, where $\tilde{R}_3 < R_3$. More in particular, the harmonic retrieval problem was solved in Reference [37] by means of a best multilinear rank-(K, K, K) approximation, where K is the number of signal poles. In this paper, we will work via an unsymmetric rank-(K, K, K') approximation, where K' is possibly smaller than K.
- In the presence of closely spaced sinusoids, the data tensor is ill conditioned in one mode. The unsymmetric tensor approximation allows us to deal with this ill conditioned mode in a proper way. We show by means of numerical simulations that our method outperforms HTLSDstack. The fact that one can treat an ill-conditioned mode in another manner than the well-conditioned modes is a profound advantage of working in the framework of multilinear algebra. Indeed, in matrix algebra column rank and row rank are equal, and one does not have the freedom to treat one mode in a different way than the other.

The paper is structured as follows. In Section 2, we summarize the HTLS and HTLSDstack techniques and point out the structure that remains unexploited. Material of this section will be further used in the tensor-based algorithm. In Section 3, we introduce the multilinear counterpart of HTLSDstack, both the theoretical foundations and the resulting algorithm. Section 4 presents some numerical simulation results. Section 5 is the conclusion.

2. SUBSPACE METHODS FOR HARMONIC RETRIEVAL: HTLS AND HTLSDstack

In this section, we briefly describe the reference methods against which our new algorithm will be compared. For a more detailed explanation we refer to [3,4,29–32].

2.1. The HTLS algorithm

Let us consider $\hat{x}_n$, the noise-free version of x_n in Equation (1). We first arrange the data samples of each channel into a $(L \times M)$ Hankel matrix as follows:

$$\boldsymbol{H} = \begin{pmatrix} x_0 & x_1 & x_2 & \cdots & x_{M-1} \\ x_1 & x_2 & \iddots & \cdots & \vdots \\ x_2 & \iddots & \iddots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & x_{N-2} \\ x_{L-1} & \cdots & \cdots & x_{N-2} & x_{N-1} \end{pmatrix} \tag{3}$$

where the column size L and the row size M are such that $N = L + M - 1$. The matrix $\boldsymbol{H}$ can be factorized as the product of three matrices as follows:

$$\boldsymbol{H} = \boldsymbol{S}\boldsymbol{C}\boldsymbol{T}^{\mathrm{T}} \tag{4}$$

 Copyright © 2009 John Wiley & Sons, Ltd.

in which

$$\boldsymbol{S} = \begin{pmatrix} 1 & \cdots & 1 \\ z_1 & \cdots & z_K \\ z_1^2 & \cdots & z_K^2 \\ \vdots & \vdots & \vdots \\ z_1^{(L-1)} & \cdots & z_K^{(L-1)} \end{pmatrix} \tag{5}$$

and

$$\boldsymbol{T} = \begin{pmatrix} 1 & \cdots & 1 \\ z_1 & \cdots & z_K \\ z_1^2 & \cdots & z_K^2 \\ \vdots & \vdots & \vdots \\ z_1^{(M-1)} & \cdots & z_K^{(M-1)} \end{pmatrix} \tag{6}$$

are respectively a $(L \times K)$ and a $(M \times K)$ Vandermonde matrix with $z_1, \ldots, z_K$ as generators, and $\boldsymbol{C} = \text{diag}\{c_1 \ldots, c_K\}$. In Equation (4), the superscript $^{\mathsf{T}}$ denotes the transpose. This type of decomposition is called a Vandermonde decomposition (VDMD). It is well known that $\boldsymbol{H}$ is a rank-K matrix when $L \geqslant K$ and $M \geqslant K$. In other words, if $L > K$ and $M > K$, then $\boldsymbol{H}$ is rank deficient and its rank corresponds to the number of poles.

Subspace methods for harmonic retrieval start from the fact that knowledge of the column space of $\boldsymbol{S}$ alone suffices to find the parameters of interest. Stack the K dominant left singular vectors of $\boldsymbol{H}$ in the $(L \times K)$ matrix $\widehat{\boldsymbol{U}}$. In the noise-free case, the columns of $\widehat{\boldsymbol{U}}$ span the same subspace as the columns of $\boldsymbol{S}$, while in the presence of noise they yield a LS approximation of the column span of $\boldsymbol{S}$. The crucial property, on which the estimation of the poles is based, is that the matrix $\boldsymbol{S}$, being Vandermonde, has a shift-invariance property that can be expressed as

$$\boldsymbol{S}_{\downarrow}\boldsymbol{Z} = \boldsymbol{S}^{\uparrow} \tag{7}$$

where the up (down) arrow placed behind a matrix stands for deleting the top (bottom) row of the considered matrix and $\boldsymbol{Z} = \text{diag}(z_1, z_2, \ldots, z_K) \in \mathbb{C}^{K \times K}$. This property implies that

$$\widehat{\boldsymbol{U}}^{\uparrow} = \widehat{\boldsymbol{U}}_{\downarrow}\widetilde{\boldsymbol{Z}} \tag{8}$$

where the $(K \times K)$ matrix $\widetilde{\boldsymbol{Z}}$ has the poles $\{z_k\}$ as its eigenvalues. In the HTLS algorithm, Equation (8) is solved in TLS sense [41]. Finally, the eigenvalues λ_k of the matrix $\widetilde{\boldsymbol{Z}}$ are an estimate of the signal poles z_k:

$$\lambda_k = \exp\{(-\hat{\alpha}_k + 2\pi j\hat{\nu}_k)\Delta t\} \tag{9}$$

where $\hat{\alpha}_k$ and $\hat{\nu}_k$ denote the estimates of α_k and ν_k, respectively.

2.2. Decimated signals: the HTLSDstack algorithm

When signal poles are very close, one could increase the accuracy by oversampling the signal. However, this leads to a large matrix $\boldsymbol{H}$. Computation of the K dominant left singular vectors of this matrix and solving Equation (8) in TLS sense may be too expensive.

On the other hand, according to the Nyquist sampling theorem one can multiply the sampling time interval Δt by a factor D as long as $\max_k |\nu_k| < \nu_s/(2D)$. Decimating (or downsampling) a signal by a factor D consists of constructing a signal that is D times shorter by picking every Dth sample or, in other words, by multiplying the sampling time interval by a factor D. We assume that N, the number of samples of $\hat{x}_n$, is an integer multiple of D. Therefore we can generate D decimated sequences of $N_D = N/D$ samples:

$$\hat{x}_n^{(d)} : \begin{cases} \hat{x}_{nD+d} = \displaystyle\sum_{k=1}^{K} c_k z_k^{nD+d} \\ n = 0, \ldots, N_D - 1 \\ d = 0, \ldots, D - 1 \end{cases} \tag{10}$$

where d is the sequence (or channel) number and n the decimated time index. Each of these sequences can be arranged in a $(L_D \times M_D)$ Hankel matrix $\boldsymbol{H}_d$, in analogy with Equation (3). The column size L_D and the row size M_D are such that $N_D = L_D + M_D - 1$. The matrix $\boldsymbol{H}_d$ can now be factorized as the product of three matrices as follows:

$$\boldsymbol{H}_d = \boldsymbol{S}\boldsymbol{C}_d\boldsymbol{T}^{\mathsf{T}} \tag{11}$$

in which $\boldsymbol{S}$ and $\boldsymbol{T}$ are respectively a $(L_D \times K)$ and a $(M_D \times K)$ Vandermonde matrix with $z_1^D, \ldots, z_K^D$ as generators, and in which

$$\boldsymbol{C}_d = \begin{pmatrix} c_1 z_1^d & & \\ & \ddots & \\ & & c_K z_K^d \end{pmatrix} \tag{12}$$

In the HTLSDstack method, the D Hankel matrices are stacked in a $(L_D \times D \cdot M_D)$ block Hankel matrix:

$$\boldsymbol{H} = \left[\boldsymbol{H}_0 \,|\, \boldsymbol{H}_1 \,|\, \cdots \,|\, \boldsymbol{H}_{D-1}\right] \tag{13}$$

This results in the following decomposition:

$$\boldsymbol{H} = \boldsymbol{S}\left[\boldsymbol{C}_0\boldsymbol{T}^{\mathsf{T}} \,|\, \boldsymbol{C}_1\boldsymbol{T}^{\mathsf{T}} \,|\, \cdots \,|\, \boldsymbol{C}_{D-1}\boldsymbol{T}^{\mathsf{T}}\right] \tag{14}$$

It appears from Equation (14) that $\boldsymbol{H}$ is a rank-K matrix for $L_D \geqslant K$ and $M_D \cdot D \geqslant K$. Stated otherwise, if $L_D > K$ and $M_D \cdot D > K$, then $\boldsymbol{H}$ is rank deficient, with the rank equal to the number of poles. Like in Section 2.1, the parameters of interest can be retrieved from the dominant K-dimensional subspace of the column space of the matrix $\boldsymbol{H}$. We stack the K dominant left singular vectors of $\boldsymbol{H}$ in a $(L_D \times K)$ matrix $\widehat{\boldsymbol{U}}$. Then we compute a $(K \times K)$ matrix $\widetilde{\boldsymbol{Z}}$ from Equation (8). Finally, the decimated signal poles z_k^D are obtained as the eigenvalues λ_k of $\widetilde{\boldsymbol{Z}}$:

$$\lambda_k = \exp\{(-\hat{\alpha}_k + 2\pi j\hat{\nu}_k)D\Delta t\} \tag{15}$$

There exist, to our knowledge, no methods with which the optimal dimensionality of the data matrix $\boldsymbol{H}$ can be computed. In Reference [32] it was deduced from experiments that the best performance is obtained for square submatrices, i.e. $L_D \approx M_D$. In the presence of noise, in order to allow a good estimation of the

Copyright © 2009 John Wiley & Sons, Ltd.

signal subspace, the column size L_D should be at least one order of magnitude greater than the model order K. Typically, in NMR this does not pose a problem: oversampling factors can range from 2 to 10 [31,32] and typical numbers of data points per spectrum are 1024 or 2048 points.

The main computational complexity of the algorithm consists of two SVDs: the SVD of the $(L_D \times M_D.D)$ data matrix in Equation (13) and the SVD to compute the TLS solution of Equation (8), involving a matrix of size $(L_D \times 2K)$. The computational aspects of these SVDs will not be discussed in detail, since there exist various algorithms, each with specific properties, that can be applied. Efficient algorithms exploit the Hankel structure and compute only a partial SVD (K singular vectors). Computational aspects are discussed in some more detail in Reference [32].

To make the link with the higher order approach in the next section, we remark that (i) the repetition of the Vandermonde matrix $\boldsymbol{T}^{\mathrm{T}}$ in each block of $\boldsymbol{H}$ in Equation (14) is not reflected by the SVD of $\boldsymbol{H}$, and that (ii) there exists an obvious relationship between the matrices $\boldsymbol{C}_d$ that has not been exploited so far:

$$\boldsymbol{C}_d = \begin{pmatrix} z_1 & & \\ & \ddots & \\ & & z_K \end{pmatrix} \boldsymbol{C}_{d-1} \tag{16}$$

which can equivalently be written as

$$\boldsymbol{C}_d = \begin{pmatrix} z_1^d & & \\ & \ddots & \\ & & z_K^d \end{pmatrix} \boldsymbol{C}_0 \tag{17}$$

with $\boldsymbol{C}_0 = \mathrm{diag}\{c_1 \ldots, c_K\}$. A multilinear approach will allow one to take more of the structure into account.

3. THE TENSOR APPROACH

3.1. Background material

First, we briefly describe the tensor-algebraic material that will be needed in the further derivations. Interested readers are referred to [42–44] for more details. We start with some basic algebra. In analogy with the matrix case, a third-order tensor $\mathcal{A}$ is called rank-1 when it equals the outer product of three vectors $U^{(1)}$, $U^{(2)}$, $U^{(3)}$:

$$a_{i_1 i_2 i_3} = u_{i_1}^{(1)} u_{i_2}^{(2)} u_{i_3}^{(3)} \tag{18}$$

for all index values, which we will denote as

$$\mathcal{A} = U^{(1)} \circ U^{(2)} \circ U^{(3)} \tag{19}$$

A decomposition of a tensor in rank-1 terms is called a parallel factor decomposition (PARAFAC) [45] or a canonical decomposition (CANDECOMP) [46]. Also see References [47,48] and the references therein. An n-mode vector of a tensor $\mathcal{A}$ is a vector obtained from $\mathcal{A}$ by varying only the nth index. As such, the n-mode vectors are the generalization of column and row vectors in the matrix case. The subspace spanned by all the n-mode vectors, for a given value of n, is called the n-mode vector space. The dimension of this subspace is called the n-mode rank of $\mathcal{A}$, denoted by $\mathrm{rank}_n(\mathcal{A})$. A difference between matrices and tensors is that, for tensors, the different n-mode ranks are not necessarily equal, while for matrices, column and row rank are always the same. A third-order tensor whose n-mode ranks are equal to R_n $(1 \leqslant n \leqslant 3)$ is called a rank-$(\mathrm{R}_1, \mathrm{R}_2, \mathrm{R}_3)$ tensor. A tensor–matrix product can be defined as follows [49]: we have

$$\mathcal{A} = \mathcal{B} \bullet_1 \boldsymbol{V}^{(1)} \bullet_2 \boldsymbol{V}^{(2)} \bullet_3 \boldsymbol{V}^{(3)} \tag{20}$$

if and only if

$$a_{i_1 i_2 i_3} = \sum_{j_1 j_2 j_3} b_{j_1 j_2 j_3} v_{j_1 i_1}^{(1)} v_{j_2 i_2}^{(2)} v_{j_3 i_3}^{(3)} \tag{21}$$

for all values of the indices.

The matrix method described in Section 2 is based on the computation of the dominant subspace of the data matrix by means of an SVD. In the remainder of this section, we explain in general terms how this problem generalizes to tensors. In References [42,49,50], an Nth-order natural extension of the matrix SVD is discussed. Let the superscript * denote the complex conjugate. For the third-order case, we have

Theorem III.1 (Third-order Singular Value Decomposition)**.** *Every complex $(I_1 \times I_2 \times I_3)$-tensor $\mathcal{A}$ can be written as the product*

$$\mathcal{A} = \mathcal{B} \bullet_1 \boldsymbol{V}^{(1)} \bullet_2 \boldsymbol{V}^{(2)} \bullet_3 \boldsymbol{V}^{(3)} \tag{22}$$

in which

- $\boldsymbol{V}^{(n)} = [V_1^{(n)} V_2^{(n)} \ldots V_{I_n}^{(n)}]$ *is a unitary $(I_n \times I_n)$-matrix,*
- *$\mathcal{B}$ is a complex all-orthogonal and ordered $(I_1 \times I_2 \times I_3)$-tensor. This means that the matrices $\boldsymbol{B}_{n,\alpha} = \mathcal{B}_{i_n=\alpha}$, obtained by fixing the nth index of $\mathcal{B}$ to α, have the following properties:*
  - *all-orthogonality: two matrices $\boldsymbol{B}_{n,\alpha}$ and $\boldsymbol{B}_{n,\beta}$ are orthogonal for all possible values of n, α and β subject to $\alpha \neq \beta$:*

$$\langle \boldsymbol{B}_{n,\alpha}, \boldsymbol{B}_{n,\beta} \rangle = \sum_{r,s} \left(\boldsymbol{B}_{n,\alpha}\right)_{rs} \left(\boldsymbol{B}_{n,\beta}\right)_{rs}^{*} = 0 \quad \textit{when} \quad \alpha \neq \beta \tag{23}$$

  - *ordering*

$$\|\boldsymbol{B}_{n,1}\| \geqslant \|\boldsymbol{B}_{n,2}\| \geqslant \cdots \|\boldsymbol{B}_{n,I_n}\| \geqslant 0 \tag{24}$$

for all possible values of n.

The Frobenius norm $\|\boldsymbol{B}_{n,i}\|$, symbolized by $\sigma_i^{(n)}$, is the ith n-mode singular value of $\mathcal{A}$ and the vector $V_i^{(n)}$ is the ith n-mode singular vector.

An important property is that the n-mode vectors corresponding to non-zero n-mode singular values form an orthonormal basis for the n-mode vector space of $\mathcal{A}$. The number of (significant) n-mode singular values equals the (numerical) n-mode rank of $\mathcal{A}$. Matrix $\boldsymbol{V}^{(n)}$ $(1 \leqslant n \leqslant 3)$ can be easily computed as the matrix of left singular vectors of a matrix in which all n-mode vectors of $\mathcal{A}$ are stacked one after the other. The n-mode singular values are the singular values of this matrix.

One way to generalize the best rank-R approximation problem in matrix algebra is as follows:

 Copyright © 2009 John Wiley & Sons, Ltd.

Given a complex third-order tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$*, find a rank-*$(R_1, R_2, R_3)$ *tensor* $\widehat{\mathcal{A}}$*, with* $rank_1(\widehat{\mathcal{A}}) \leqslant R_1$*,* $rank_2(\widehat{\mathcal{A}}) \leqslant R_2$ *and* $rank_3(\widehat{\mathcal{A}}) \leqslant R_3$*, that minimizes the least squares cost function*

$$f\left(\widehat{\mathcal{A}}\right) = \left\|\mathcal{A} - \widehat{\mathcal{A}}\right\|^2 = \sum_{i_1,i_2,i_3} \left|a_{i_1,i_2,i_3} - \hat{a}_{i_1,i_2,i_3}\right|^2 \tag{25}$$

Due to the *n*-rank constraints, $\widehat{\mathcal{A}}$ can be decomposed as

$$\widehat{\mathcal{A}} = \mathcal{B} \bullet_1 \boldsymbol{V}^{(1)} \bullet_2 \boldsymbol{V}^{(2)} \bullet_3 \boldsymbol{V}^{(3)} \tag{26}$$

in which $\boldsymbol{V}^{(1)} \in \mathbb{C}^{I_1 \times R_1}$, $\boldsymbol{V}^{(2)} \in \mathbb{C}^{I_2 \times R_2}$, $\boldsymbol{V}^{(3)} \in \mathbb{C}^{I_3 \times R_3}$ each have orthonormal columns and $\mathcal{B} \in \mathbb{C}^{R_1 \times R_2 \times R_3}$ is an all-orthogonal tensor. As is well known, in the matrix case, the best rank-R approximation can simply be obtained by truncation of the matrix SVD. An important difference between matrices and higher order tensors, however, is that the best rank-(R_1, R_2, R_3) approximation cannot in general be obtained by mere truncation of the HOSVD. Nevertheless, due to the ordering constraint (24), the truncated HOSVD is usually a pretty good approximation, albeit not the optimal one. The best rank-(R_1, R_2, R_3) approximation can be calculated by means of tensor generalizations of algorithms for the calculation of a dominant subspace. A higher order orthogonal iteration (HOOI) method has been studied in References [43,51,52]. The HOOI algorithm is of the alternating least squares (ALSs) type and its convergence rate is linear. Recently, methods with a higher convergence rate have been developed [53–56].

3.2. Higher order Vandermonde structure of decimated data sequences

The concepts introduced in Section 3.1 can be used to derive a tensor-based algorithm for harmonic retrieval as follows. We form a $(L_D \times M_D \times D)$ tensor $\mathcal{H}$ by stacking the D Hankel matrices $\boldsymbol{H}_d$ one behind the other. In other words, each $(L_D \times M_D)$ slice of $\mathcal{H}$ is a Hankel matrix that corresponds to an artificially created channel, obtained by decimation. This is visualized in Figure 1. An element $h_{i_1 i_2 i_3}$ of this tensor is given by

$$h_{i_1 i_2 i_3} = x_{(i_1-1)D+(i_2-1)D+(i_3-1)} \tag{27}$$

where $1 \leqslant i_1 \leqslant L_D, 1 \leqslant i_2 \leqslant M_D, 1 \leqslant i_3 \leqslant D, L_D + M_D - 1 = N_D$ or $(L_D + M_D - 1)D = N$. As long as the latter constraint is verified, the dimensions of the tensor may be chosen by the user. Let the one-dimensional noise-free complex signal $\dot{x}_n$ be a K-pole time domain signal modelled by Equation (10). In what follows, we assume that $L_D > K$. If we replace each sample in Figure 1 by this model, an element of the tensor can be expressed as follows:

$$h_{i_1 i_2 i_3} = \sum_{k=1}^{K} c_k \left(z_k^{(i_1-1)D} \cdot z_k^{(i_2-1)D} \cdot z_k^{i_3-1}\right) \tag{28}$$

Hence, the tensor $\mathcal{H}$ is a weighted sum of third-order rank-1 tensors, consisting of the outer product of three Vandermonde vectors

$$\mathcal{H} = \sum_{k=1}^{K} c_k \cdot \begin{pmatrix} 1 \\ \left(z_k^D\right)^1 \\ \left(z_k^D\right)^2 \\ \vdots \\ \left(z_k^D\right)^{L_D-1} \end{pmatrix} \circ \begin{pmatrix} 1 \\ \left(z_k^D\right)^1 \\ \left(z_k^D\right)^2 \\ \vdots \\ \left(z_k^D\right)^{M_D-1} \end{pmatrix} \circ \begin{pmatrix} 1 \\ z_k^1 \\ z_k^2 \\ \vdots \\ z_k^{D-1} \end{pmatrix} \tag{29}$$

Equation (29) is a PARAFAC decomposition of $\mathcal{H}$. In principle, the poles could be obtained by computing this decomposition. In this paper, however, we will develop a subspace-type method that allows us to manipulate the modes of $\mathcal{H}$ in an unsymmetric

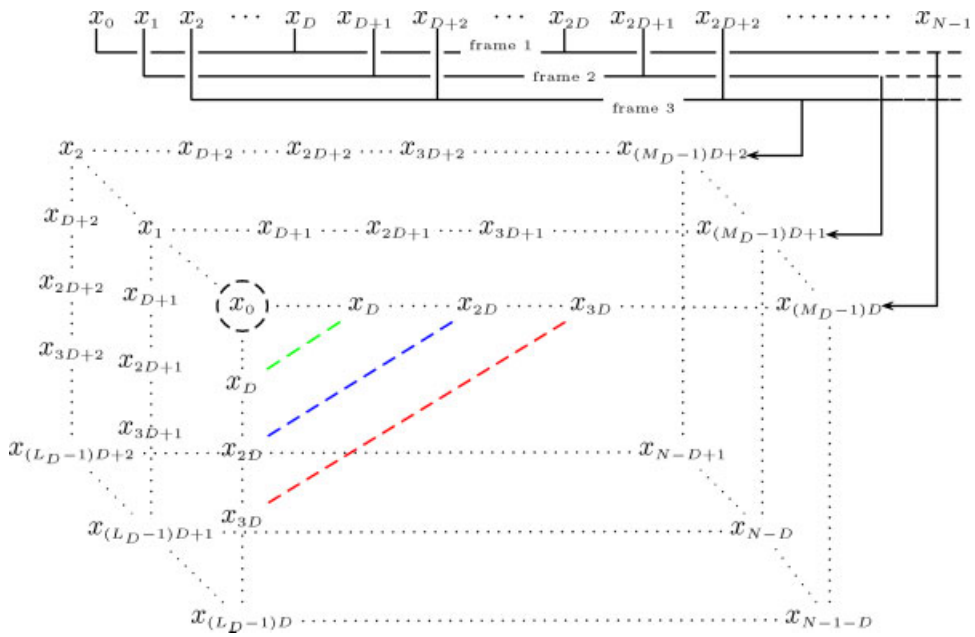


Figure 1. Construction of a tensor from Hankel matrices representing a decimated signal. The dotted lines delimit the tensor while the dashed lines show its partial Hankel structure. This figure is available in colour online at www.interscience.wiley.com/journal/cem

Copyright © 2009 John Wiley & Sons, Ltd.

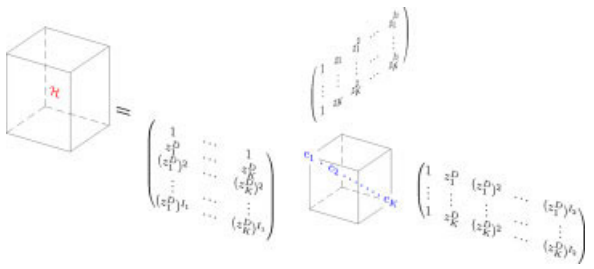


Figure 2. Visualization of the HOVDMD of $\mathcal{H}$. This representation is equivalent to the PARAFAC decomposition (29) for $I_1 = L_D - 1$, $I_2 = M_D - 1$ and $I_3 = D - 1$. This figure is available in colour online at www.interscience.wiley.com/journal/cem

way. It will turn out that this is advantageous when signal poles are close. Equation (29) can also be written as

$$\mathcal{H} = \mathcal{C} \bullet_1 \boldsymbol{S}^{(1)} \bullet_2 \boldsymbol{S}^{(2)} \bullet_3 \boldsymbol{S}^{(3)} \tag{30}$$

in which $\mathcal{C}$ is a diagonal $(K \times K \times K)$ core tensor that contains the K complex amplitudes c_k. The matrices $\boldsymbol{S}^{(1)} \in \mathbb{C}^{K \times L_D}$, $\boldsymbol{S}^{(2)} \in \mathbb{C}^{K \times M_D}$ and $\boldsymbol{S}^{(3)} \in \mathbb{C}^{K \times D}$ are Vandermonde, with generators $\{z_k^D\}$, $\{z_k^D\}$ and $\{z_k\}$, respectively. In analogy with Equation (4), decomposition (30) is called a higher order Vandermonde decomposition (HOVDMD). It is visualized in Figure 2.

3.3. Pole extraction

Notice that the 1-mode vectors of $\mathcal{H}$ are generated as linear combinations of the Vandermonde vectors that form the columns of $\boldsymbol{S}^{(1)}$ in Equation (30). On the other hand, the 1-mode vectors are also generated as linear combinations of the K 1-mode singular vectors of $\mathcal{H}$ that correspond to the non-zero 1-mode singular values. This suggests that we can work in analogy with the matrix case. We should now compute the K-dimensional dominant subspace of the 1-mode space of tensor $\mathcal{H}$, and then exploit shift-invariance to extract the poles. Let the part of the HOSVD of $\mathcal{H}$ that corresponds to the non-zero part of its core tensor be given by

$$\mathcal{H} = \widetilde{\mathcal{B}} \bullet_1 \widetilde{\boldsymbol{V}}^{(1)} \bullet_2 \widetilde{\boldsymbol{V}}^{(2)} \bullet_3 \widetilde{\boldsymbol{V}}^{(3)} \tag{31}$$

in which $\widetilde{\mathcal{B}}$ is a complex $(K \times \min(K, M_D) \times \min(K, D))$ all-orthogonal core tensor, and in which $\widetilde{\boldsymbol{V}}^{(1)}$, $\widetilde{\boldsymbol{V}}^{(2)}$ and $\widetilde{\boldsymbol{V}}^{(3)}$ are a $(L_D \times K)$, $(M_D \times \min(K, M_D))$ and $(D \times \min(K, D))$ column-wise orthonormal matrix, respectively. We now have that

$$\widetilde{\boldsymbol{V}}^{(1)} = \boldsymbol{S}^{(1)} \boldsymbol{Q} \tag{32}$$

in which $\boldsymbol{Q}$ is a $(K \times K)$ non-singular matrix. One can now proceed in the same way as in Section 2 to compute the signal poles from $\widetilde{\boldsymbol{V}}^{(1)}$. However, one comment is in order. The derivation so far assumed the absence of noise. If noise is present, tensor $\mathcal{H}$ will not be exactly rank-$(K, \min(K, M_D), \min(K, D))$. As explained in Section 3.1, the best rank-$(K, \min(K, M_D), \min(K, D))$ approximation of $\mathcal{H}$ and hence its dominant 1-mode subspace cannot, in general, be obtained by mere truncation of the HOSVD. Instead we compute the best rank-$(K, \min(K, M_D), \min(K, D))$ approximation

$$\widehat{\mathcal{H}} = \widehat{\mathcal{B}} \bullet_1 \widehat{\boldsymbol{U}}^{(1)} \bullet_2 \widehat{\boldsymbol{U}}^{(2)} \bullet_3 \widehat{\boldsymbol{U}}^{(3)} \tag{33}$$

in which $\widehat{\mathcal{B}}$ is a complex $(K \times \min(K, M_D) \times \min(K, D))$ all-orthogonal core tensor, and in which $\widehat{\boldsymbol{U}}^{(1)}$, $\widehat{\boldsymbol{U}}^{(2)}$ and $\widehat{\boldsymbol{U}}^{(3)}$ are a $(L_D \times K)$, $(M_D \times \min(K, M_D))$ and $(D \times \min(K, D))$ column-wise orthonormal matrix, respectively. We will estimate the signal poles from $\widehat{\boldsymbol{U}}^{(1)}$.

Note that in this tensor method, it was the structure of the matrices $\boldsymbol{C}_d$ in Equation (12) that caused the rank-R structure of tensor $\mathcal{H}$. Moreover, we determine two matrices $\widehat{\boldsymbol{U}}^{(2)}$ and $\widehat{\boldsymbol{U}}^{(3)}$, which is one matrix more than in the matrix case. This corresponds to the fact that the matrix $\boldsymbol{T}^{\mathrm{T}}$ is repeated in each block in Equation (14). More precisely, the matrix $\widehat{\boldsymbol{U}}^{(2)}$ acts on the space associated with $\boldsymbol{T}$, while $\widehat{\boldsymbol{U}}^{(3)}$ acts on the space associated with the matrices $\boldsymbol{C}_d$.

Like for HTLSDstack it is difficult to make hard statements on which choice of the dimensions L_D and M_D is optimal. Numerical experiments indicate that good results are obtained for $L_D \approx M_D$, like in the matrix case.

3.4. Unsymmetric tensor approximation

In the preceding section, we showed that it is natural to compute the signal poles from the best rank-$(K, \min(K, M_D), \min(K, D))$ approximation of $\mathcal{H}$. However, only the mode-1 K-dimensional dominant subspace is of interest. In this case, the following theorem, whose proof is given in the appendix, shows that it is not necessary to confine ourselves to a rank-$(K, \min(K, M_D), \min(K, D))$ approximation of tensor $\mathcal{H}$:

Theorem III.2 (Unsymmetric tensor approximation)**.** *Consider a tensor* $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$ *that is rank-*$(\mathrm{R}_1, \mathrm{R}_2, \mathrm{R}_3)$*. Let the HOSVD of* $\mathcal{A}$ *be given by*

$$\mathcal{A} = \mathcal{B} \bullet_1 \boldsymbol{V}^{(1)} \bullet_2 \boldsymbol{V}^{(2)} \bullet_3 \boldsymbol{V}^{(3)}$$

Then the best rank-$(\mathrm{R}_1, \mathrm{R}_2, \tilde{\mathrm{R}}_3)$ *approximation of* $\mathcal{A}$*, with* $\tilde{\mathrm{R}}_3 < \mathrm{R}_3$*, is obtained by truncation of* $\mathcal{B}$ *and* $\boldsymbol{V}^{(3)}$*.*

Taking a value for $\tilde{\mathrm{R}}_3$ that is strictly lower than R_3 is, particularly, useful when the smallest mode-3 singular values of $\mathcal{A}$ are in the order of magnitude of the noise level. It is then numerically preferable to extract the dominant $\tilde{\mathrm{R}}_3$-dimensional mode-3 vector space.

Let us turn back to the problem. The decimative approach is most often used in the case where poles are very close. This means that, in Equation (30), the Vandermonde matrix $\boldsymbol{S}^{(3)}$ is very ill-conditioned because its generators are very close in the complex z-plane. The matrices $\boldsymbol{S}^{(1)}$ and $\boldsymbol{S}^{(2)}$ are less ill-conditioned because of the decimation: the generators $\{z_k^D\}$ are less close than the generators $\{z_k\}$ when the damping factors are not too big. The condition of the problem can be easily evaluated by inspecting the mode-n singular values in the HOSVD of $\mathcal{H}$. If, for instance, there is a large gap between the 3-mode singular values, it is better to compute the best rank-$(K, \min(K, M_D), K')$ approximation of $\mathcal{H}$ with $K' < \min(K, D)$. Concentrating on the dominant part of $\mathcal{H}$ increases the robustness.

3.5. Tensor-based algorithm for harmonic retrieval using decimated data sequences

In this section, we summarize the tensor-based algorithm for the estimation of the K signal poles of an oversampled signal x_n. This

 Copyright © 2009 John Wiley & Sons, Ltd.

algorithm is called HO-HTLSDstack, which stands for higher order Hankel TLS method for decimated sequences stacked in a tensor.
Input: complex signal x_n,
Output: the set of signal poles z_k.

- decimate x_n making sure that no aliasing occurs and construct D signals $\hat{x}_n^{(d)}$ (see Equation (10)),
- map $\hat{x}_n^{(d)}$ to a $(L_D \times M_D \times D)$-tensor $\mathcal{H}$ as in Figure 1 (see Equation (27)),
- find the best rank-$(K, \min(K, M_D), K')$ approximation of $\mathcal{H}$ with $K' \leq \min(K, D)$, and let the columns of $\widehat{\boldsymbol{U}}^{(1)}$ form a basis for the K-dimensional dominant subspace of the 1-mode vector space,
- substitute $\widehat{\boldsymbol{U}}^{(1)}$ for $\widehat{\boldsymbol{U}}$ in the overdetermined set of linear Equations (8), and compute the TLS solution $\widetilde{\boldsymbol{Z}}$,
- compute the eigenvalues of $\widetilde{\boldsymbol{Z}}$ and derive the parameter estimates (Equation (15)).

In the third step, the HOOI algorithm can be used [43]. It is an iteration that alternates between (i) the computation of the dominant K-dimensional subspace of the column space of a $(L_D \times KK')$ matrix, (ii) the computation of the dominant K-dimensional subspace of the column space of a $(M_D \times KK')$ matrix, (iii) the computation of the dominant K'-dimensional subspace of the column space of a $(D \times K^2)$ matrix. (We assume that $K < M_D$.) In each step, one can start from the estimate of the dominant subspace obtained in the previous step. Truncation of the HOSVD of $\mathcal{H}$, involving the computation of the dominant K-dimensional subspace of the column space of a $(M_D \times L_D D)$ matrix and the computation of the dominant K'-dimensional subspace of the column space of a $(D \times L_D M_D)$ matrix, usually yields a good initial value for the HOOI algorithm. The key point is that the matrices in the iteration steps are always small in one dimension. This guarantees that the overall complexity stays fairly low compared to non-decimative methods, which involve the computation of the dominant K-dimensional subspace of a $(DL_D \times DM_D)$ matrix. Furthermore, after a couple of HOOI steps the convergence can be speed up as explained in References [53–56].

4. RESULTS

The simulations below were done on a processor AMD DURON 2.8Ghz and using Matlab R14 as software.

4.1. Two-peak, undamped signal

This example is taken from Reference [32]. The signal consists of two undamped complex exponentials ($K = 2$) contaminated by complex circularly symmetric WGN. The true parameters are: $\nu_1 = 0.2$ Hz, $\nu_2 = 0.205$ Hz, $a_1 = a_2 = 1$, $\alpha_1 = \alpha_2 = 0$, $\varphi_1 = \varphi_2 = 0$, $\Delta t = 0.1$ and $N = 1000$. The normalized frequencies are 0.02 and 0.0205, respectively, which allows for a decimation factor $D = 10$ without risk of aliasing. A Monte-Carlo simulation consisting of 100 independent runs is carried out. The results shown in Figure 3 are formulated in terms of a relative root mean square error (RRMSE), defined as follows:

$$\text{RRMSE}(\nu_k) = \sqrt{\frac{\sum_{i=1}^{N\text{runs}} (\hat{\nu}_{k,i} - \nu_k)^2}{N\text{runs}}} \cdot \frac{100}{0.5 \times |\nu_2 - \nu_1|} \tag{34}$$

where $\hat{\nu}_{k,i}$ is the estimate of ν_k at run i and Nruns is the number of trials. For instance, RRMSE$(\nu_1) < 100\%$ means that the frequency ν_1 is on the average in the interval $[\nu_1 - |\nu_2 - \nu_1|/2, \nu_1 + |\nu_2 - \nu_1|/2]$. Note that this value does not mean that the estimates are not consistent since the RRMSE is an average value. The measure is only a way to compare the accuracy of the results to $|\nu_2 - \nu_1|/2$; the value 100% can be considered as a conservative threshold. For the matrix approach, the parameters are computed from the dominant column subspace $\widehat{\boldsymbol{U}} \in \mathbb{C}^{50\times 2}$ of the matrix $\boldsymbol{H} \in \mathbb{C}^{50\times 500}$ (see Equation (14)). As far as the tensor approach is concerned, we computed a best rank-(2, 2, 2) and a best rank-(2, 2, 1) approximation of the third-order tensor $\mathcal{H} \in \mathbb{C}^{50\times 50\times 10}$. Then we used the 1-mode dominant subspace $\widehat{\boldsymbol{U}}^{(1)} \in \mathbb{C}^{50\times 2}$ to estimate the parameters. We used square submatrices which leads to the best performance of HTLSDstack [32]. From Figure 3 it is clear that for high signal-to-noise ratios (SNRs) the tensor approach and the matrix approach perform nearly the same. However, the tensor approach is more robust when the SNR decreases. Moreover, Figure 3 shows that the best rank-(2, 2, 2) approximation is less reliable than the best rank-(2, 2, 1) approximation. The reason is that the mode-3 subspace is ill-conditioned.

4.2. Two-peak, damped signal

We consider two closely spaced peaks whose parameters are: $\nu_1 = 0.2, \nu_2 = 0.22, a_1 = a_2 = 1, \alpha_1 = 0.01, \alpha_2 = 0.02, \varphi_1 = \varphi_2 = 0$, $\Delta t = 0.04$ and $N = 625$. The decimation factor $D = 25$. The Monte-Carlo simulation consists of 100 trials. In this example, we

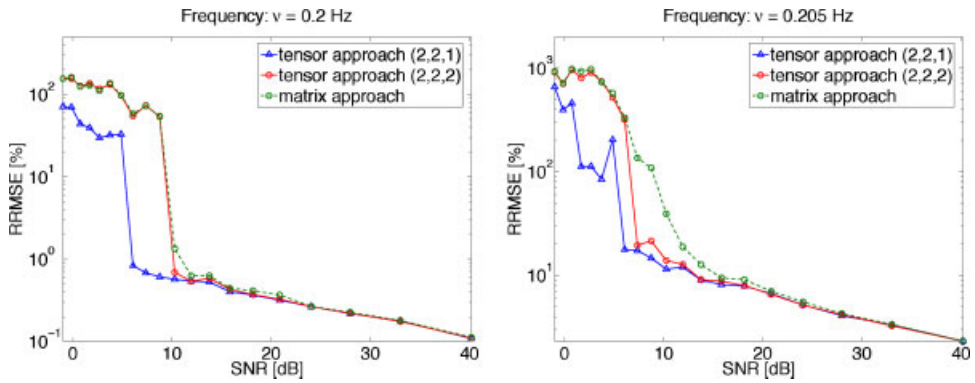


Figure 3. Comparison of the tensor and the matrix approach for the two-peak, undamped example. This figure is available in colour online at www.interscience.wiley.com/journal/cem

Copyright © 2009 John Wiley & Sons, Ltd.

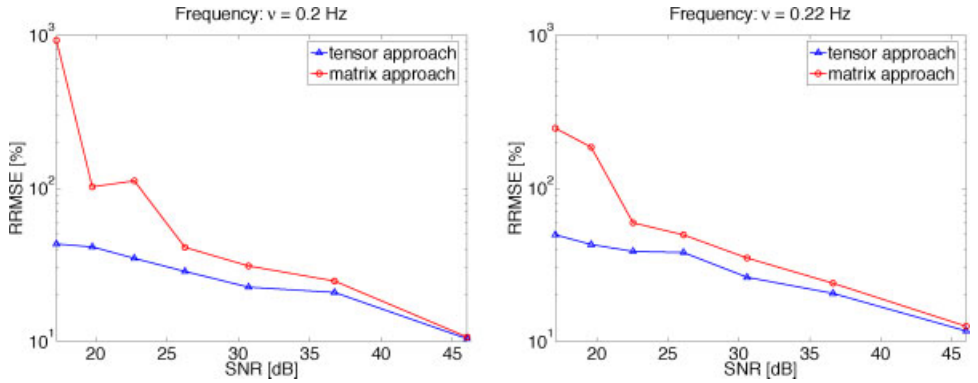


Figure 4. Comparison of the tensor and the matrix approach for the two-peak, damped example. This figure is available in colour online at www.interscience.wiley.com/journal/cem

used the best rank-2 approximation of the (13×325) matrix $\boldsymbol{H}$ and the best rank-$(2, 2, 1)$ approximation of the $(13 \times 13 \times 25)$-tensor $\mathcal{H}$. The results, shown in Figure 4, show that the tensor approach is clearly more accurate than the matrix approach when the SNR is lower than 45 dB. Above this value, the improvement given by the tensor approach is not significant.

4.3. Five-peak, damped signal

In this example, we simulate a typical ^{31}P NMR signal of perfused rat liver. This signal has been used several times as a test case for comparing the performance of algorithms in the literature of NMR spectroscopy, see e.g. Reference [4], where the test case

Table I. Set of parameters for the five-peak example

Peak k	ν_k [Hz]	α_k [s^{-1}]	a_k [a.u.][a]	φ_k [°][b]
1	−1379	208	6.1	15
2	−685	256	9.9	15
3	−271	197	6.0	15
4	353	117	2.8	15
5	478	808	17.0	15

[a] a.u., arbitrary units.
[b] $(\varphi_k \times \pi/180)$ expresses the phase in radians.

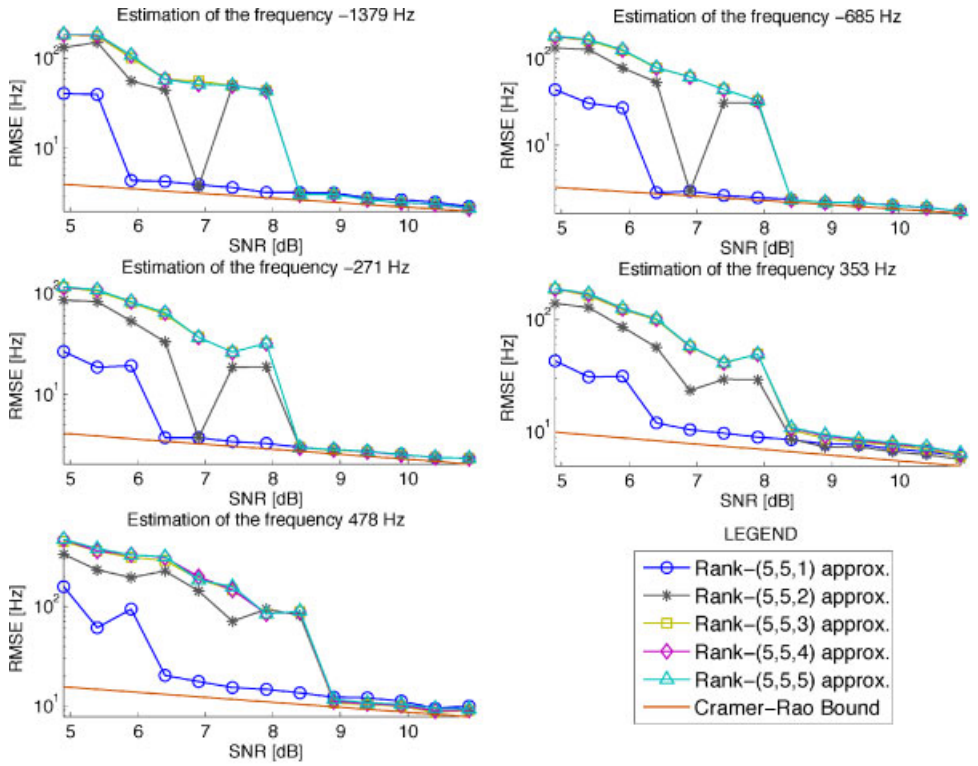


Figure 5. Influence of the choice of parameter K' in HO-HTLSDstack for the example in Section 4.3. This figure is available in colour online at www.interscience.wiley.com/journal/cem

 Copyright © 2009 John Wiley & Sons, Ltd.

was first introduced. The signal represents in a simplified way the NMR response of the ATP molecule, where the α-triplet and the β and γ doublets have been reduced to single Lorentzian peaks. The damped exponential model as taken here, or equivalently the Lorentzian model in the frequency domain, is the model that is the most often used in *in vivo* NMR spectroscopy. The exact parameters of the signal are reported in Table I. The signal is generated at a sampling rate of 50 KHz and is composed of 500 samples. Peak 5 has a very large damping factor and therefore has a very broad spectral content. As a consequence it overlaps with peak 4 and makes the frequency separation very difficult. The performance of the algorithms will be expressed in terms of the RMSE to allow for a clear comparison with the Cramér-Rao (CR) bound. Here, the RMSE is expressed as follows:

$$\mathrm{RMSE}(\nu_k) = \sqrt{\frac{\sum_{i=1}^{Nruns}(\hat{\nu}_{k,i} - \nu_k)^2}{Nruns}} \tag{35}$$

The Monte-Carlo simulation consists of 500 trials. The sampling rate allows for a downsampling of the signal by a factor $D = 10$. Therefore 10 signals of 50 samples have been generated and each of them has been stacked in a (26×25)-Hankel matrix. Hence, we process a $(26 \times 25 \times 10)$ data tensor. In Figure 5, we show the effect of choosing different values for the parameter K' in the HO-HTLSDstack algorithm. For an SNR above 9 dB all curves coincide. However, the threshold below which the algorithm becomes inaccurate can be decreased by choosing a smaller value for K'. As such, the best results were obtained for the best rank-(5, 5, 1) approximation of $\mathcal{H}$.

In Figure 6, the performance of our new algorithm is compared to HTLSDstack and HTLS. HO-HTLSDstack and HTLS diverge from the CR bound at the same noise level while HTLSDstack diverges for a higher SNR. The latter method is less robust against noise but, on the other hand, has the lowest computational complexity. It should be mentioned that HTLSDstack already successfully compares to most competing decimation-based algorithms [32].

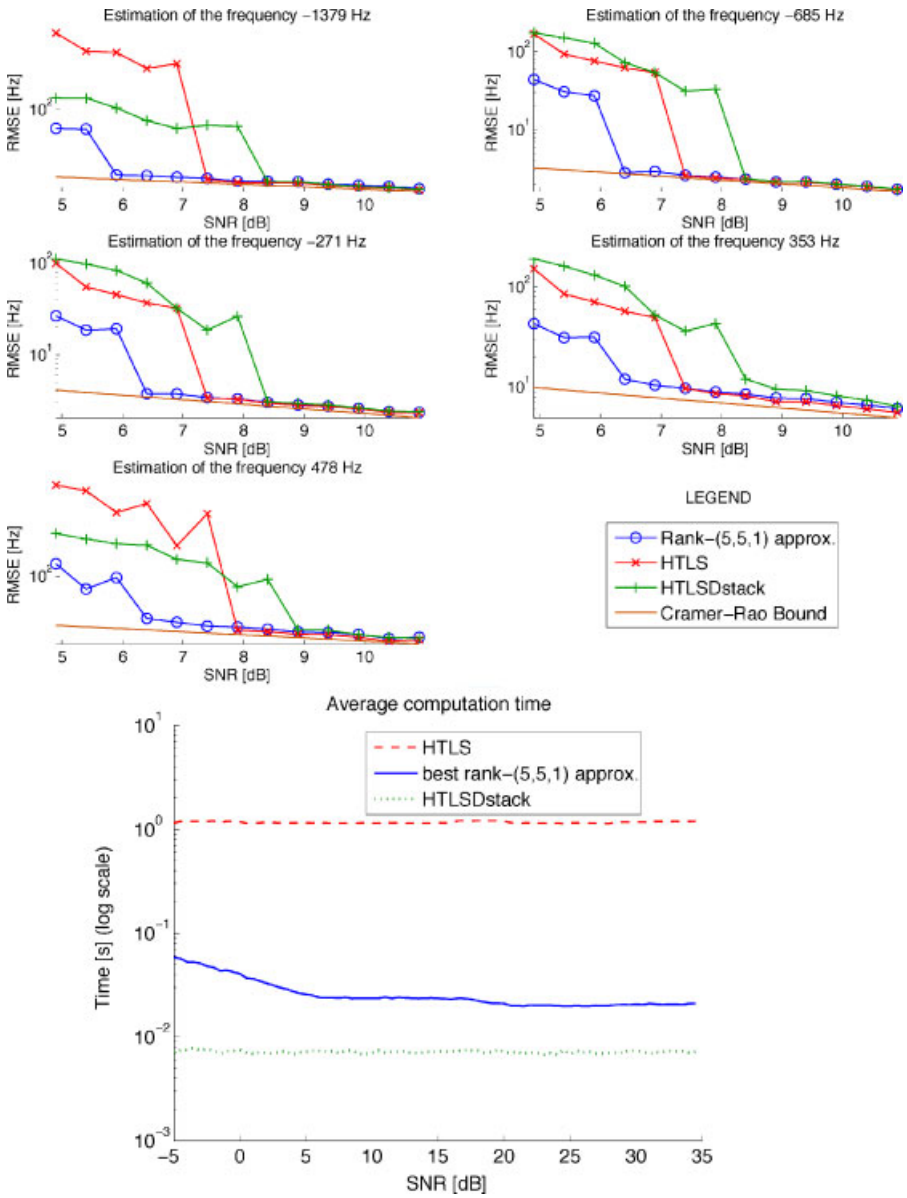


Figure 6. Comparison in accuracy and time of HO-HTLSDstack to the decimation-based matrix algorithm HTLSDstack and to the non-decimative matrix algorithm HTLS (example in Section 4.3). This figure is available in colour online at www.interscience.wiley.com/journal/cem

Copyright © 2009 John Wiley & Sons, Ltd.

The price to pay for the higher robustness of HO-HTLSDstack is a slight augmentation of the computation cost, compared to HTLSDstack. However, this cost is still more than an order of magnitude lower than that of the non-decimative HTLS method.

5. CONCLUSION

In this paper, we have considered harmonic analysis of a signal whose poles are potentially very close. When the signal can be downsampled, it is possible to use so-called decimative methods. Currently, the most effective matrix-based technique is the HTLSDstack algorithm. Starting from this approach, we have shown that we can stack the data in a third-order tensor instead of a matrix. The tensor formalism allows one to take into account a structure in the third mode of the tensor that is not exploited in the matrix-based algorithms. We have derived a higher order version of HTLSDstack. Ill-conditioned modes can be treated in a different manner than well-conditioned modes. We have demonstrated through numerical examples that the new tensor technique is more robust than its matrix counterpart.

Acknowledgments

A large part of this research was carried out when L. De Lathauwer was with the French Centre National de la Recherche Scientifique (CNRS). The authors thank T. McKelvey (Chalmers University of Technology, Göteborg, Sweden) for helpful discussions. The research has been supported by Research Council KUL: GOA-AMBioRICS, CoE EF/05/006 Optimization in Engineering, IDO 05/010 EEG-fMRI, Campus Impuls Financiering (CIF1), STRT1/08/023; Flemish Government: FWO: G.0360.05 (EEG, Epileptic), FWO-G.0321.06 (Tensors/Spectral Analysis); Belgian Federal Science Policy Office IUAP P6/04 (DYSCO, 'Dynamical systems, control and optimization', 2007–2011); EU: FAST (FP6-MC-RTN-035801), Neuromath (COST-BM0601).

REFERENCES

1. Chen H. Subspace-based parameter estimation of exponentially damped sinusoids with application to nuclear magnetic resonance spectroscopy data. *PhD thesis*, Department of Electrical Engineering, Katholieke Universiteit Leuven, 1996.
2. Pels P. Analysis and improvement of quantification algorithms for magnetic resonance spectroscopy. *PhD thesis*, Katholieke Universiteit Leuven, Belgium, January 2005.
3. Van Huffel S. Enhanced resolution based on minimum variance estimation and exponential data modeling. *Signal Processing* 1993; **33**(3): 333–355.
4. Van Huffel S, Chen H, Decanniere C, Van Hecke P. Algorithm for time-domain NMR data fitting based on total least squares. *J. Magn. Reson. A* 1994; **110**: 228–237.
5. Vanhamme L, in't Zandt HJA, Van Huffel S, Van Hecke P. Biomedical magnetic resonance spectroscopic quantitation: a review of modern time-domain analysis methods. In *Proceedings of the 12th ProRISC Workshop on Circuits, Systems and Signal Processing (ProRISC)*, Veldhoven, November 2001; 1–8.
6. Vanhamme L, Sundin T, Van Hecke P, Van Huffel S. MR spectroscopy quantitation: a review of time-domain methods. *NMR Biomed.* 2001; **14**(4): 233–246.
7. Vanhamme L, Van Huffel S. Multichannel quantification of biomedical magnetic resonance spectroscopy signals. In *Proceedings of SPIE, Advanced Signal Processing Algorithms, Architectures, and Implementations VIII*, Luk FT (ed.), Vol. 3461, San Diego, California, July 1998; 237–248.
8. Boyer R, Abed-Meraim K. Efficient parametric modeling for audio transients. In *Proceedings of the 5th International Conference on Digital Audio Effects (DAFx-02)*, Hamburg, Germany, September 2002. Online article: http//:www.dfax.de/
9. Boyer R, Abed-Meraim K. Audio modeling based on delayed sinusoids. *IEEE Trans. Speech Audio Process.* 2004; **12**(2): 110–120.
10. Nieuwenhuijse J, Heusdens R, Deprettere EF. Robust exponential modeling of audio signal. In *Proceedings of International Conference on Acoustics, Speech and Signal Processing*, volume 6, 1998.
11. Verhelst W, Hermus K, Lemmerling P, Wambacq P, Van Huffel S. Modeling audio of damped sinusoids using total least squares algorithms. In *Total Least Squares and Errors-in-Variables Modeling: Analysis, Algorithms and Applications*, Van Huffel S, Lemmerling P (eds). Kluwer Academic Publishers: Dordrecht, 2002; 331–340.
12. Jensen J. Sinusoidal models for speech signal processing. *PhD thesis*, Aalborg University, Denmark, 2000.
13. Lemmerling P, Dologlou I, Van Huffel S. Speech compression based on exact modeling and structured total least norm optimization. In *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP98)*, Seattle, USA, May 1998; 353–356.
14. Elad M, Milanfar P, Golub GH. Shape from moments—an estimation theory perspective. *IEEE Trans Signal Process* 2004; **52**: 1814–1829.
15. Rippert L. Optical fiber for damage monitoring in carbon fiber reinforced plastic composite materials. *PhD thesis*, Katholieke Universiteit Leuven, March 2003.
16. Bresler Y, Macovski A. Exact maximum likelihood parameter estimation of superimposed exponential signals in noise. *IEEE Trans. Acoust.* 1986; **34**: 1081–1089.
17. Stoica P, Sharmen KC. Maximum likelihood methods for direction-of-arrival estimation. *IEEE Trans. Acoust.* 1990; **38**(7): 1132–1143.
18. Kumaresan R, Tufts DW. Estimating the parameters of exponentially damped sinusoids and pole-zero modelling in noise. *IEEE Trans. Acoust* 1982; **30**(6): 833–840.
19. Kumaresan R, Scharf LL, Shaw AK. An algorithm for pole-zero modeling and spectral analysis. *IEEE Trans. Acoust.* 1986; **34**(6): 637–640.
20. Bhaskar Rao DV, Arun KS. Model-based processing of signals: a state-space approach. *Proc. IEEE* 1992; **80**: 283–309.
21. Kung S-Y, Arun KS, Bhaskar Rao DV. State-space and singular value decomposition-based approximation methods for the harmonic retrieval problem. *J. Opt. Soc. Am.* 1983; **73**(12): 1799–1811.
22. Markovsky I, Vaccaro RJ, Van Huffel S. System identification by optimal subspace estimation. *Technical Report 06–162*, Deptartment EE, K.U.Leuven, 2006. ftp://ftp.esat.kuleuven.ac.be/pub/SISTA/markovsky/reports/06-210.ps.gz
23. Hua Y, Sarkar TK. Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise. *IEEE Trans. Acoust.* 1990; **38**: 814–824.
24. Papadopoulos CK, Nikias CL. Parameter estimation of exponentially damped sinusoids using higher-order statistics. *IEEE Trans. Acoust.* 1990; **38**(8): 1424–1436.
25. Pisarenko V. The retrieval from harmonics from a covariance function. *Geophys. J. R. Astron. Soc.* 1973; **33**: 347–366.
26. Roy R, Paulraj A, Kailath T. ESPRIT—a subspace rotation approach for to estimation of parameters of cisoids in noise. *IEEE Trans. Acoust.* 1986; **34**(4): 1340–1342.
27. Roy R, Kailath T. ESPRIT—estimation of signal parameters via rotational invariance techniques. *IEEE Trans. Acoust.* 1989; **37**(7): 984–995.
28. Barkhuijsen H, de Beer R, van Ormondt D. Improved algorithm for noniterative time-domain model fitting to exponentially damped magnetic resonance signals. *J. Magn. Reson.* 1987; **73**: 553–557.
29. Chen H, Van Huffel S, Vandewalle J. Bandpass prefiltering for exponential data fitting with known frequency region of interest. *Signal Processing* 1996; **48**: 135–154.
30. Vanhamme L. Advanced time-domain methods for nuclear magnetic resonance spectroscopy data analysis. *PhD thesis*, Katholieke Universiteit Leuven, Leuven, Belgium, November 1999.
31. Morren G. Advanced signal processing applied to in-vivo spectroscopy and heart rate variability. *PhD thesis*, Faculty of Engineering, K.U.Leuven, Leuven, Belgium, May 2004.
32. Morren G, Lemmerling P, Van Huffel S. Decimative subspace-based parameter estimation technique. *Signal Processing* 2003; **83**(5): 1025–1033.

 Copyright © 2009 John Wiley & Sons, Ltd.

33. Stoica P, Nordsjö AE. Subspace-based frequency estimation in the presence of moving-average noise using decimation. *Signal Processing* 1997; **63**: 221–220.
34. Halder B, Kailath T. Efficient estimation of closely spaced sinusoidal frequencies using subspace-based methods. *IEEE Signal Process. Lett.* 1997; **4**(2): 49–51.
35. Van Huffel S, Wang Y, Vanhamme L, Van Hecke P. Automatic frequency alignment and quantitation of single resonances spectra via complex principal component analysis. *J. Magn. Reson.* 2002; **158**: 1–14.
36. Wang Y, Van Huffel S, Vanhamme L, Mastronardi N, Van Hecke P. Advanced signal processing methods for quantitation of resonances in magnetic resonance spectra. In *Proceedings of the Thirteenth IEEE Symposium on Computer-Based Medical Systems (CBMS2000)*, Houston, USA, June 2000; 63–68.
37. Papy J-M, De Lathauwer L, Van Huffel S. Exponential data fitting using multilinear algebra: the single and the multi-channel case. *Numer. Linear Algebra Appl.* 2005; **12**(8): 1025–1033.
38. Jiang T, Sidiropoulos ND, ten Berge JMF. Almost sure identifiability of multidimensional harmonic retrieval. *IEEE Trans. Signal Process.* 2001; **49**(9): 1849–1859.
39. Sidiropoulos ND. Generalizing Caratheodory's uniqueness of harmonic parameterization to *n* dimensions. *IEEE Trans. Inf. Theory* 2001; **47**(4): 1687–1690.
40. Boyer R, De Lathauwer L, Abed-Meraim K. Higher-order tensor based method for delayed exponential modeling. *IEEE Trans. Signal Process.* 2007; **55**(6): 2795–2809.
41. Van Huffel S, Vandewalle J. *The Total Least Squares Problem: Computational Aspects and Analysis, volume 9 of Frontiers in Applied Mathematics*. SIAM: Philadelphia, PA, 1991.
42. De Lathauwer L, De Moor B, Vandewalle J. A multilinear singular value decomposition. *SIAM J. Matrix Anal. Appl.* 2000; **21**(4): 1253–1278.
43. De Lathauwer L, De Moor B, Vandewalle J. On the best rank-1 and rank-(R_1, R_2,..., R_N) approximation of higher-order tensors. *SIAM J. Matrix Anal. Appl.* 2000; **21**(4): 1324–1342.
44. De Lathauwer L, Vandewalle J. Dimensionality reduction in higher-order signal processing and rank-(R_1, R_2, ..., R_N) reduction in multilinear algebra. *Linear Algebra and its Applications* 2004; **391**: 31–55. Special Issue on Linear Algebra in Signal and Image Processing.
45. Harshman RA. Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-mode factor analysis. *UCLA Working Papers in Phonetics* 1970; **16**: 1–84 (University Microfilms, Ann Arbor, Michigan, No. 10,085).
46. Carroll JD, Chang J. Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. *Psychometrika* 1970; **35**(3): 283–319.
47. De Lathauwer L. A link between the canonical decomposition in multilinear algebra and simultaneous matrix diagonalization. *SIAM J. Matrix Anal. Appl.* 2006; **28**(3): 642–666.
48. Kruskal JB. Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear Algebra Appl.* 1977; **18**(2): 95–138.
49. Tucker LR. Some mathematical notes on three-mode factor analysis. *Psychometrika* 1966; **31**: 279–311.
50. Tucker LR. The extension of factor analysis to three-dimensional matrices. In *Contributions to Mathematical Psychology*, Harris CW (ed.) Univeristy of Wisconsin Press: Madison. 1964; 109–127.
51. Kroonenberg PM. *Applied Multiway Data Analysis*. Wiley Series in Probability and Statistics, John Wiley & Sons: Chichester, 2008.
52. Kroonenberg PM, de Leeuw J. Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika* 1980; **45**: 69–97.
53. Eldén L, Savas B. A Newton–Grassmann method for computing the best multi-linear rank-(r_1, r_2, r_3) approximation of a tensor. *Technical Report LITH-MAT-R-2007-6-SE*, Department of Mathematics, Linköpings Universitet, 2007.
54. Ishteva M, De Lathauwer L, Absil P-A, Van Huffel S. Differential-geometric Newton method for the best rank-(R_1, R_2, R_3) approximation of tensors. *Numerical Algorithms*, DOI:10.1007/s11075-008-9251-2.
55. Ishteva M, De Lathauwer L, Absil P-A, Van Huffel S. Dimensionality reduction for higher-order tensors: Algorithms and applications. *Int. J. Pure Appl. Math.* 2008; **42**(3): 337–343.
56. Savas B, Lim L-H. Best multilinear rank approximation of tensors with quasi-newton methods on grassmannians. *Technical Report LITH-MAT-R-2008-01-SE*, Department of Mathematics, Linköpings Universitet, 2008.

APPENDIX: PROOF OF THEOREM III.2

Let the best rank-(R_1, R_2, $\tilde{R}_3$) approximation of $\mathcal{A}$ be equal to $\tilde{\mathcal{A}}$, and let its HOSVD be given by

$$\tilde{\mathcal{A}} = \tilde{\mathcal{B}} \bullet_1 \tilde{\boldsymbol{V}}^{(1)} \bullet_2 \tilde{\boldsymbol{V}}^{(2)} \bullet_3 \tilde{\boldsymbol{V}}^{(3)} \tag{36}$$

By definition, the unitary matrices $\tilde{\boldsymbol{V}}^{(1)}$, $\tilde{\boldsymbol{V}}^{(2)}$, $\tilde{\boldsymbol{V}}^{(3)}$ maximize [43]:

$$g\left(\tilde{\boldsymbol{V}}^{(1)}, \tilde{\boldsymbol{V}}^{(2)}, \tilde{\boldsymbol{V}}^{(3)}\right) = \left\| \mathcal{A} \bullet_1 \tilde{\boldsymbol{V}}^{(1)^{\mathrm{H}}} \bullet_2 \tilde{\boldsymbol{V}}^{(2)^{\mathrm{H}}} \bullet_3 \tilde{\boldsymbol{V}}^{(3)^{\mathrm{H}}} \right\|^2 \tag{37}$$

$$= \left\| \left(\mathcal{B} \bullet_1 \boldsymbol{V}^{(1)} \bullet_2 \boldsymbol{V}^{(2)} \bullet_3 \boldsymbol{V}^{(3)}\right) \bullet_1 \tilde{\boldsymbol{V}}^{(1)^{\mathrm{H}}} \bullet_2 \tilde{\boldsymbol{V}}^{(2)^{\mathrm{H}}} \bullet_3 \tilde{\boldsymbol{V}}^{(3)^{\mathrm{H}}} \right\|^2 \tag{38}$$

$$= \left\| \mathcal{B} \bullet_1 \left(\tilde{\boldsymbol{V}}^{(1)^{\mathrm{H}}} \cdot \boldsymbol{V}^{(1)}\right) \bullet_2 \left(\tilde{\boldsymbol{V}}^{(2)^{\mathrm{H}}} \cdot \boldsymbol{V}^{(2)}\right) \bullet_3 \left(\tilde{\boldsymbol{V}}^{(3)^{\mathrm{H}}} \cdot \boldsymbol{V}^{(3)}\right) \right\|^2 \tag{39}$$

in which the superscript $^{\mathrm{H}}$ denotes the complex conjugated transpose. For any given choice of $\boldsymbol{V}^{(3)}$, the function g reaches its maximal value for $\tilde{\boldsymbol{V}}^{(1)} = \boldsymbol{V}^{(1)}$, $\tilde{\boldsymbol{V}}^{(2)} = \boldsymbol{V}^{(2)}$. Substitution in Equation (36) yields

$$\max_{\tilde{\boldsymbol{V}}^{(1)}, \tilde{\boldsymbol{V}}^{(2)}} g\left(\tilde{\boldsymbol{V}}^{(1)}, \tilde{\boldsymbol{V}}^{(2)}, \tilde{\boldsymbol{V}}^{(3)}\right) = \left\| \mathcal{B} \bullet_3 \left(\tilde{\boldsymbol{V}}^{(3)^{\mathrm{H}}} \cdot \boldsymbol{V}^{(3)}\right) \right\|^2 \tag{40}$$

$$= \left\| \left(\mathcal{B} \bullet_3 \boldsymbol{V}^{(3)}\right) \bullet_3 \tilde{\boldsymbol{V}}^{(3)^{\mathrm{H}}} \right\|^2 \tag{41}$$

In Equation (41), $\mathcal{B} \bullet_3 \boldsymbol{V}^{(3)}$ is itself an HOSVD. From the properties of the HOSVD [42, Proposition 9, p. 1267], it follows that Equation (41) is maximal when $\tilde{\boldsymbol{V}}^{(3)}$ is taken equal to the left ($I_3 \times \tilde{R}_3$) submatrix of $\boldsymbol{V}^{(3)}$. In that case we have

$$\tilde{\boldsymbol{V}}^{(3)^{\mathrm{H}}} \cdot \boldsymbol{V}^{(3)} = \left[\boldsymbol{I}_{\tilde{R}_3 \times \tilde{R}_3} \quad \boldsymbol{0}_{\tilde{R}_3 \times (R_3 - \tilde{R}_3)}\right] \tag{42}$$

in which $\boldsymbol{I}_{\tilde{R}_3 \times \tilde{R}_3}$ and $\boldsymbol{0}_{\tilde{R}_3 \times (R_3 - \tilde{R}_3)}$ are the ($\tilde{R}_3 \times \tilde{R}_3$) identity matrix and the ($\tilde{R}_3 \times (R_3 - \tilde{R}_3)$) zero matrix, respectively. The corresponding core tensor $\tilde{\mathcal{B}}$ is obtained from

$$\tilde{\mathcal{B}} = \mathcal{A} \bullet_1 \tilde{\boldsymbol{V}}^{(1)^H} \bullet_2 \tilde{\boldsymbol{V}}^{(2)^H} \bullet_3 \tilde{\boldsymbol{V}}^{(3)^H} \tag{43}$$

$$= \mathcal{B} \bullet_1 \left(\tilde{\boldsymbol{V}}^{(1)^H} \cdot \mathbf{V}^{(1)}\right) \bullet_2 \left(\tilde{\boldsymbol{V}}^{(2)^H} \cdot \boldsymbol{V}^{(2)}\right) \bullet_3 \left(\tilde{\boldsymbol{V}}^{(3)^H} \cdot \mathbf{U}^{(3)}\right) \tag{44}$$

$$= \mathcal{B} \bullet_3 \left[\boldsymbol{I}_{\tilde{R}_3 \times \tilde{R}_3} \quad \boldsymbol{0}_{\tilde{R}_3 \times (R_3 - \tilde{R}_3)}\right] \tag{45}$$

i.e., it is simply obtained by truncation of $\mathcal{B}$. ■

Copyright © 2009 John Wiley & Sons, Ltd.